Capstone Project

# Airbnb Bookings Analysis


Presenter Name :

Vikas panchal

Naveen Kumar Batta

Manish Gupta

Md Shakib


# Content :

- Introduction.
- General overview  of dataset.
- Missing value Handling.
- Univarirate Analysis.
- Multivariate Analysis.
- Conclusion.

# Introduction :

Since 2008, guests and hosts have used Airbnb to expand on traveling possibilities and present a more unique, personalized way of experiencing the world.

Today, Airbnb became one of a kind service that is used and recognized by the whole world.

Data analysis on millions of listings provided through Airbnb is a crucial factor for the company.

These millions of listings generate a lot of data - data that can be analyzed and used for security, business decisions, understanding of customers' and providers' (hosts) behavior and performance on the platform, guiding marketing initiatives, implementation of innovative additional services and much more.

# Agenda:

- Top 25 most common words used in listing names
- Top 25 Used Words for Listing Names
- Neighbourhood group Frequency.
- Top 15 highest listing Neighbourhood.
- Total count of each room type as per listing .
- Minimum no’s of nights stayed  and preferred rooms
- Value count of Neighbourhood.
- Area vs Number of reviews

- Room types and their relation with availability in different neighbourhood groups?(contd.)
- Price vs Number of Reviews
- Find total no. nights spend as per location
- Challenges Faced
- We defined some points which can help Airbnb in their business
- Conclusion

# General overview of dataset.

- This dataset has around is mix between categorical and numeric values.
- Price is a dependent column.
- Total 16 columns are present in the dataset.
- Total observations are 48895.
- Min of Price variable is 0, max is 10000$.
- Mean price is 152$
- On an average people stay 7 days in a room.
- 75 Percentage of times minimum nights stayed is 5.
- Mean reviews given to Room/apartment is 23.

# Understanding the Data :

- There are 49,000 observations with various types of field in our dataset.
- List of field:

- Id
- Name
- Host_id
- Host_name
- Neighbourhood_group
- Neighbourhood
- Latitude
- Longitude
- Room_type
- Price
- Minimum_nights
- Number_of_reviews
- Last_review
- Reviews_per_month
- Calculated_host_listing_count
- availabilty_365

# Missing value Handling :

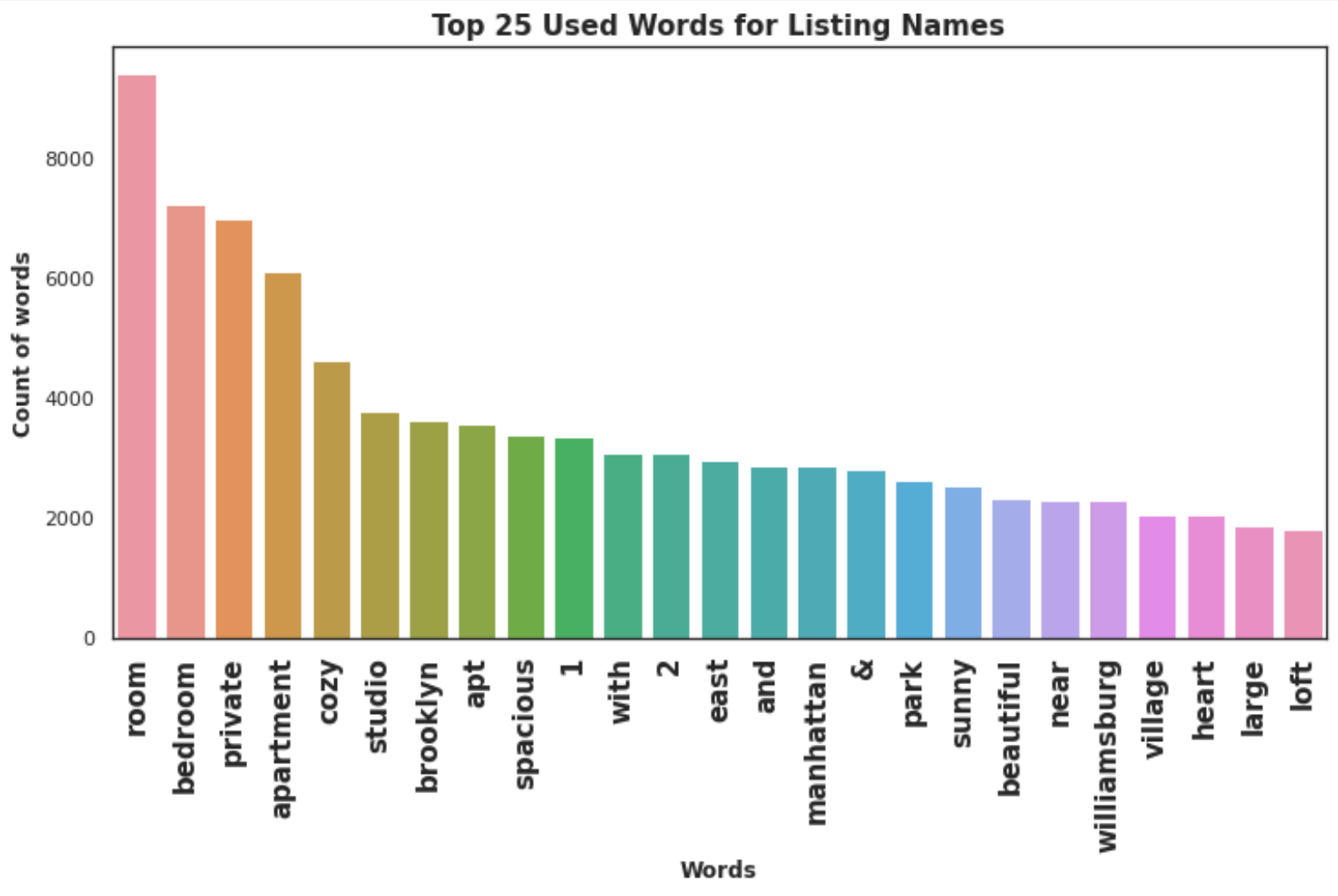
- ‘name’ = 16 null values.
- ‘host_name’ = 21
- ‘last_review’ = 10052
- ‘reviews_per_month’ = 10052
- ‘name’ column we will replaced the ‘nan’ values with corresponding ‘room_type’ values.
- ‘host_name’ will not use as those are names of individuals.
- 'reviews_per_month' we will replace with 0 for NA values.
- In ‘last_review’ We will convert its data type to catogorical and replace 'NA' with 'never'.

# Top 25 most common words used in listing names :

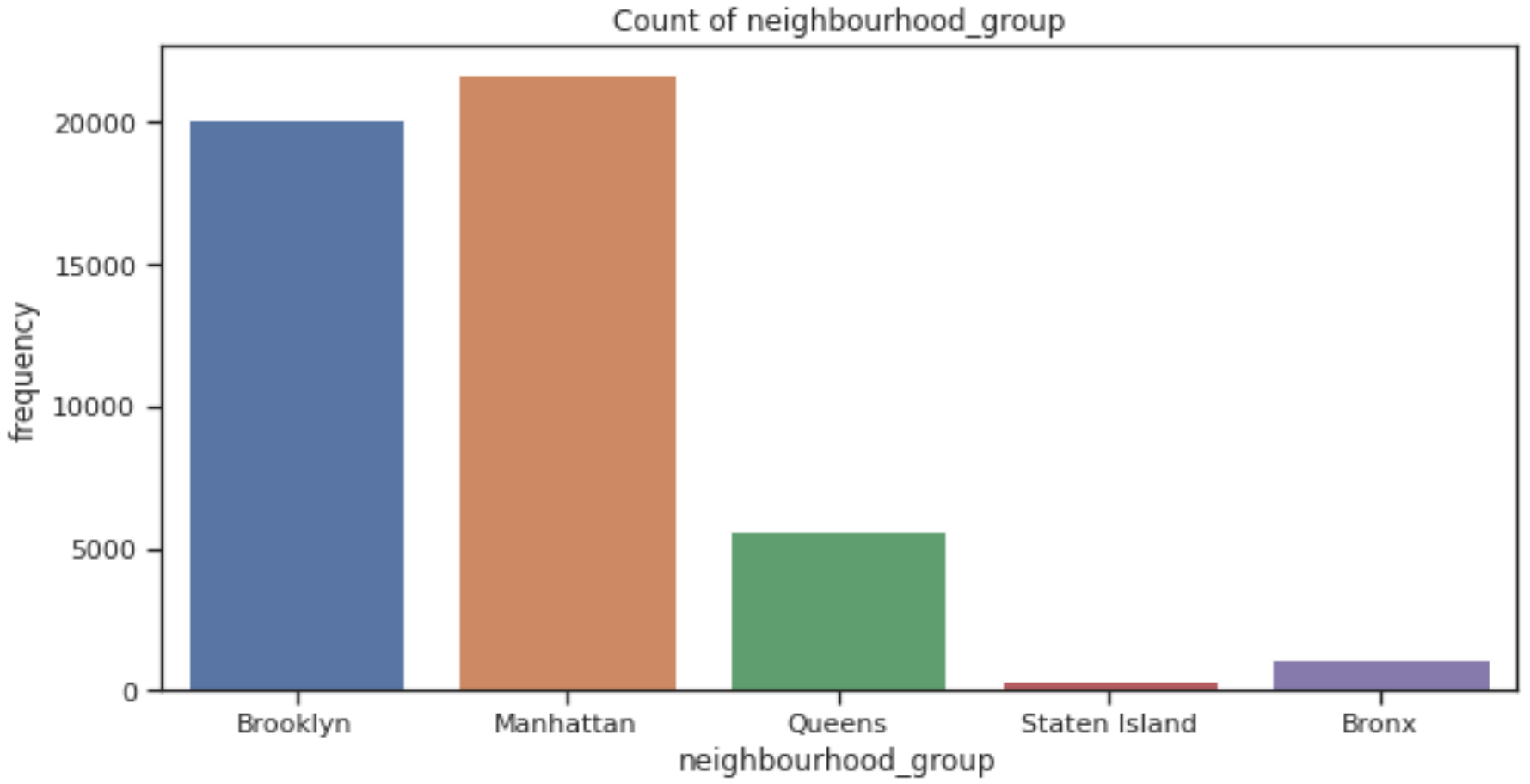
- As we can see most of the listing names include words related to property type such as ‘bedroom’, ‘cozy’, ‘private’, ‘apartment’ and ‘spacious’.

- It is interesting to see that words related to proximity or connection to public places such as ‘park’, ‘near’, ‘village’ and ‘heart’ rank lower in chart

# Top 25 Used Words for Listing Names:

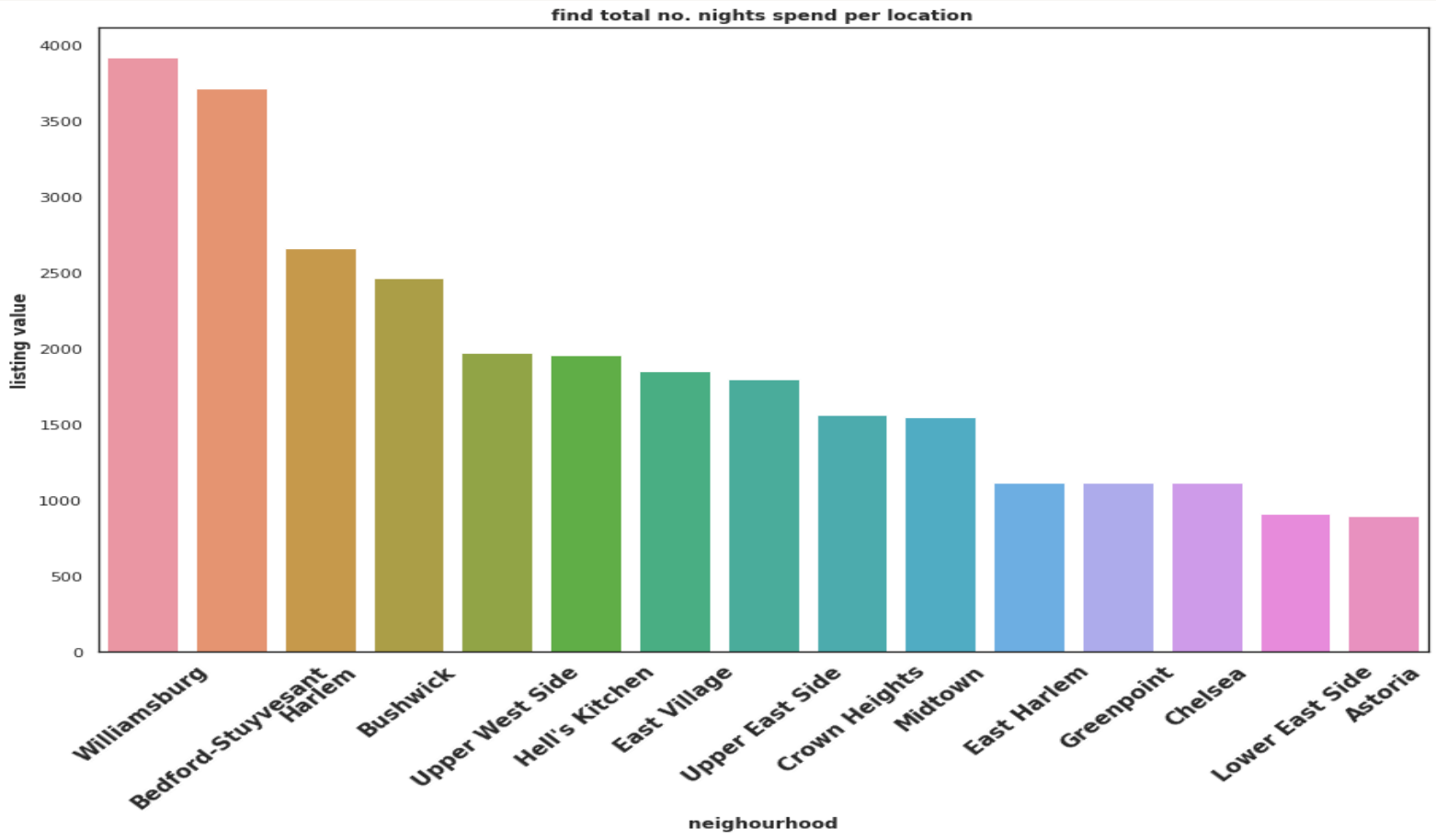


# Neighbourhood group Frequency :



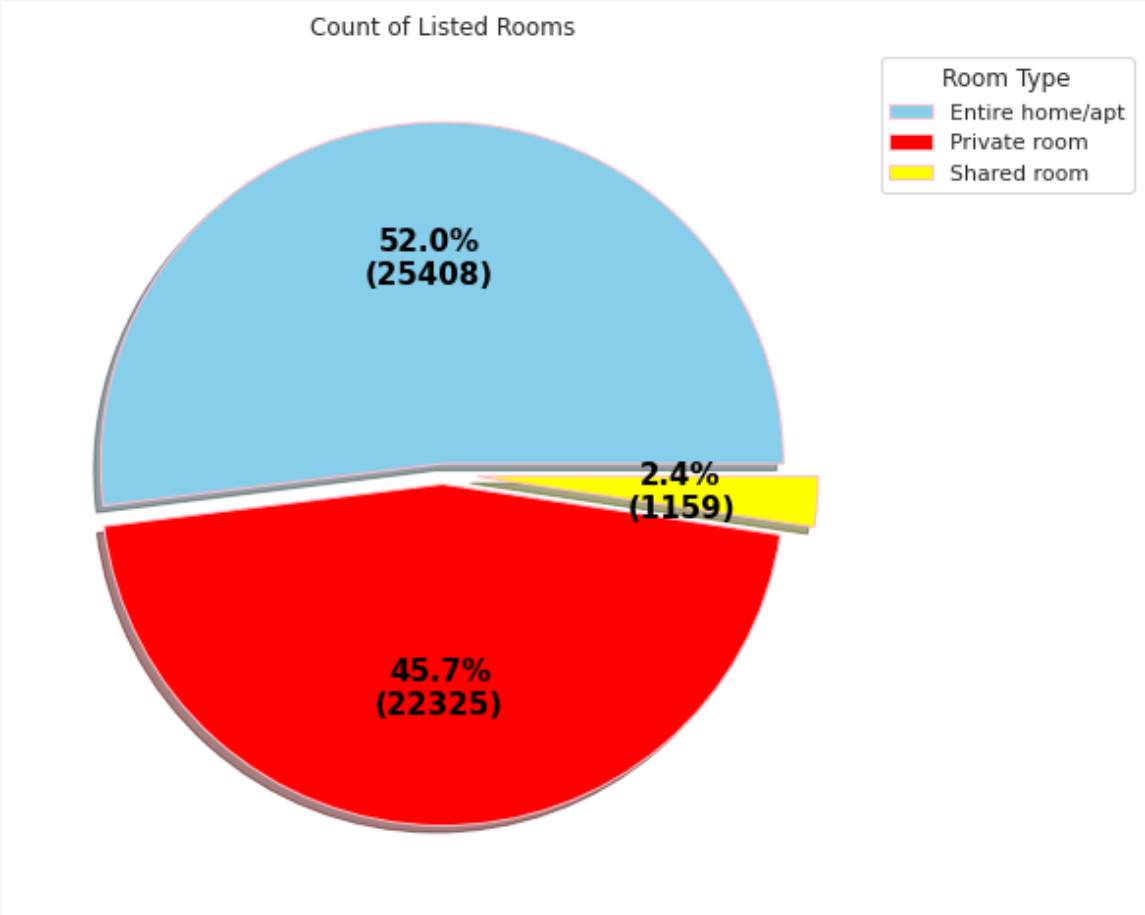
# Top 15 highest listing Neighbourhood :

- Here we can state that Williamsburg has the highest listing.
- And almost same in Bedford-Stuyvesant.
- it's a tourist attraction hence the number of listings could be highe

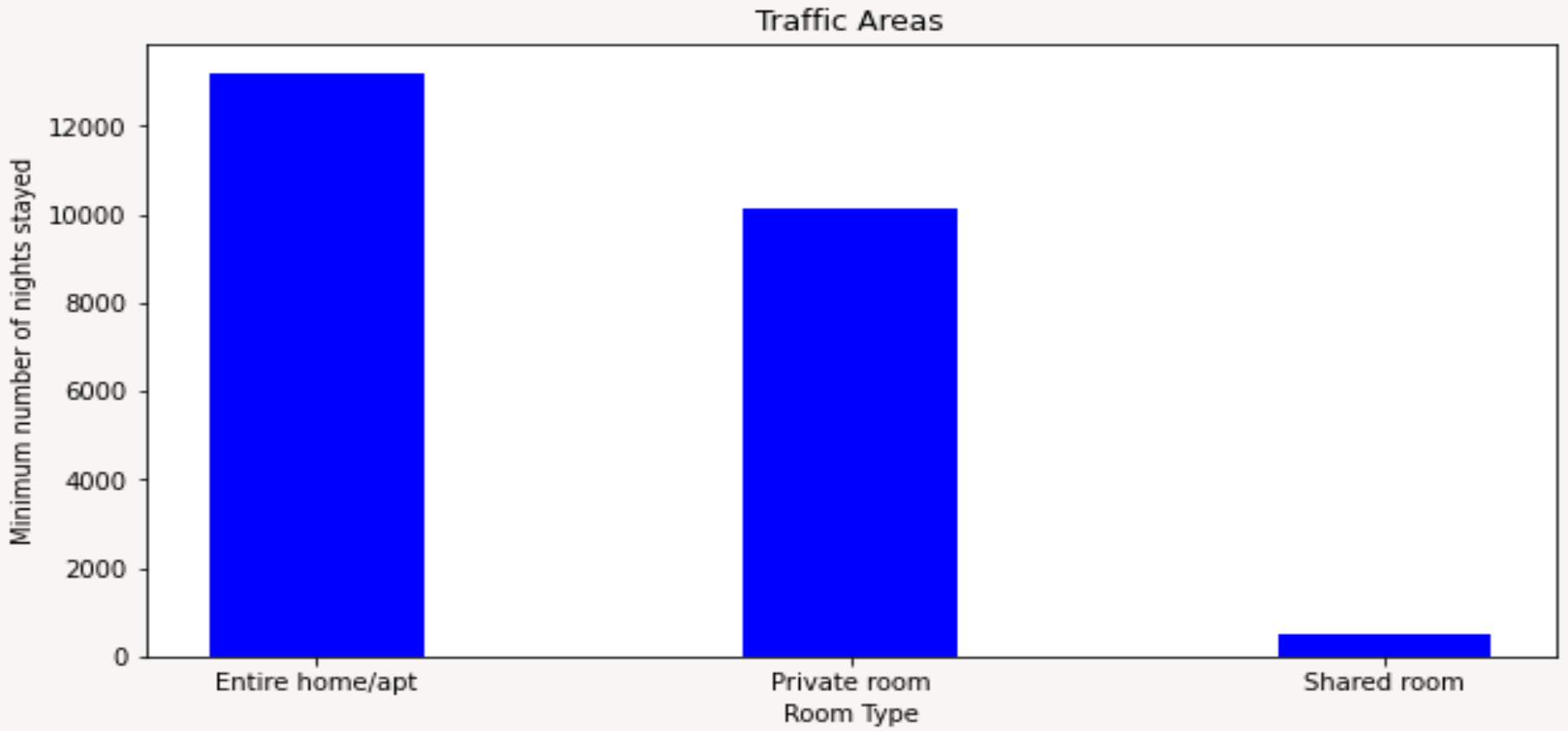


# Total count of each room type as per listing :

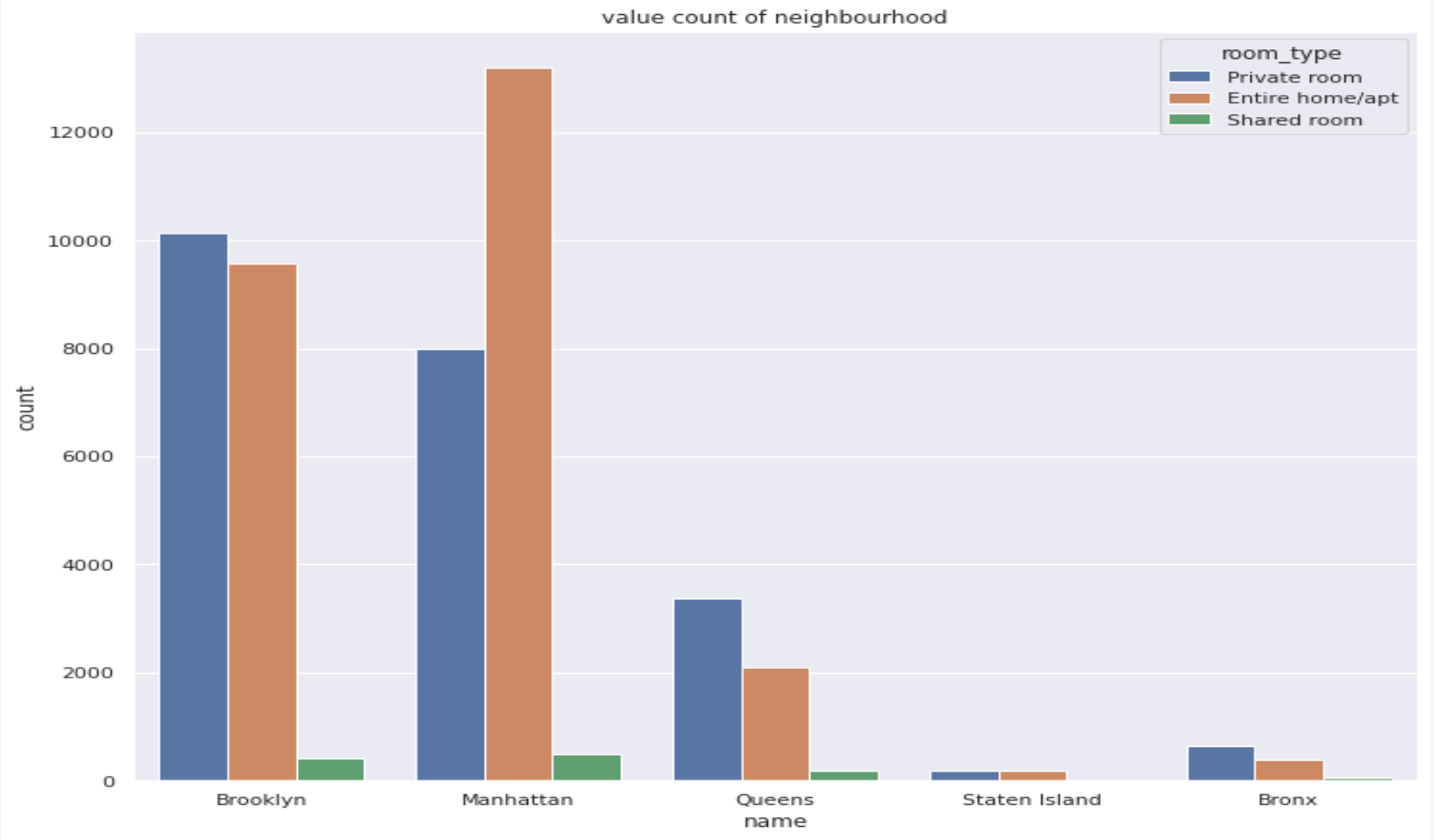
- Entire home/apt has the highest number of listing of 52% among other room.
- Private room has 45.7% of listing among other room types.
- Shared Room is the least listed room type at only 2.4% in total.



# Minimum no’s of nights stayed and preferred rooms :



# value count of Neighbourhood :



# Value count of Neighbourhood :

## Observations :

1.There are three types of rooms

2. Namely:

1.Private room

2.Entire home/apt room type

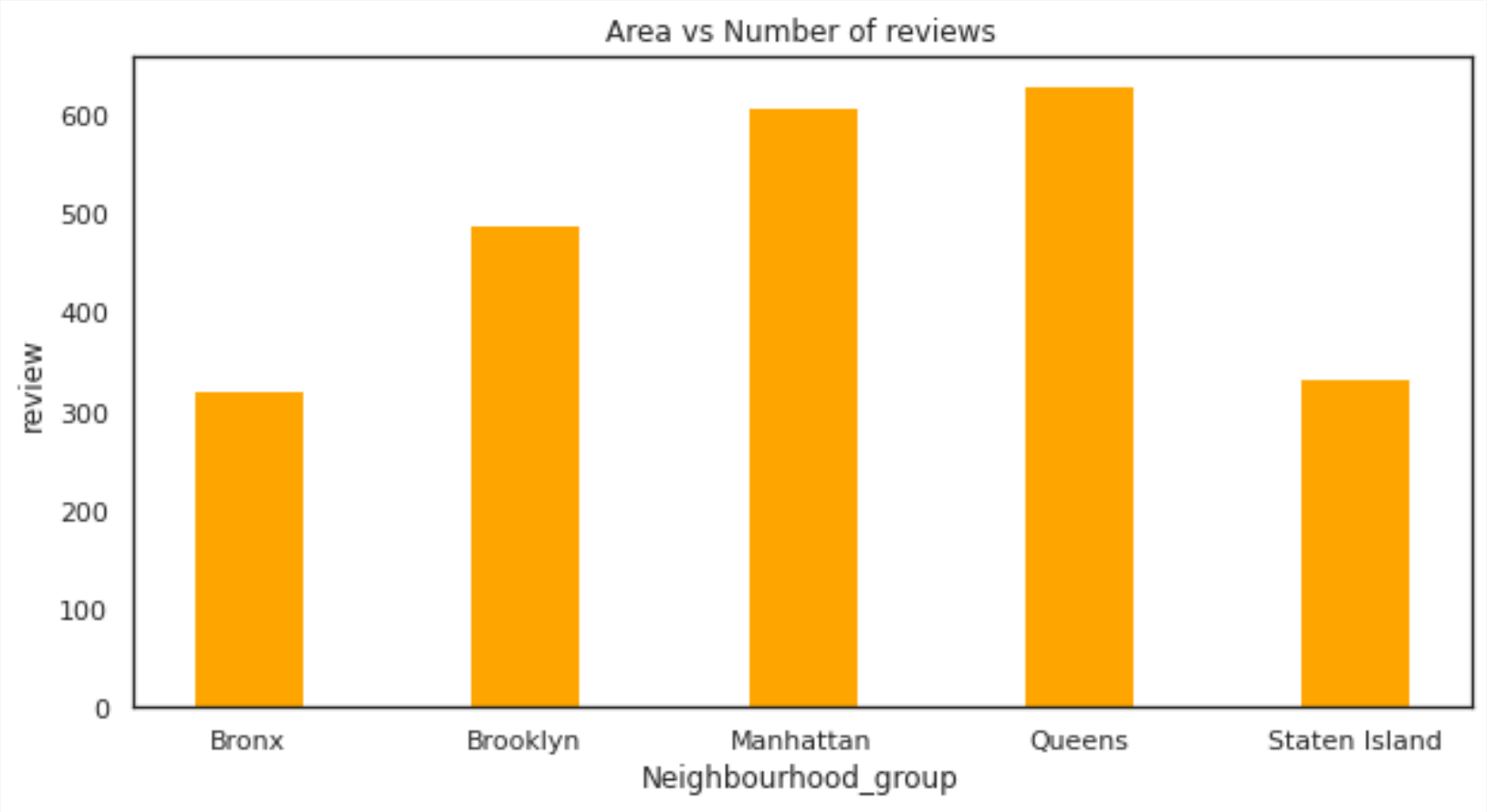
3.Shared room.

3.People mostly preferred to take whole apartment on rent followed by Private room.

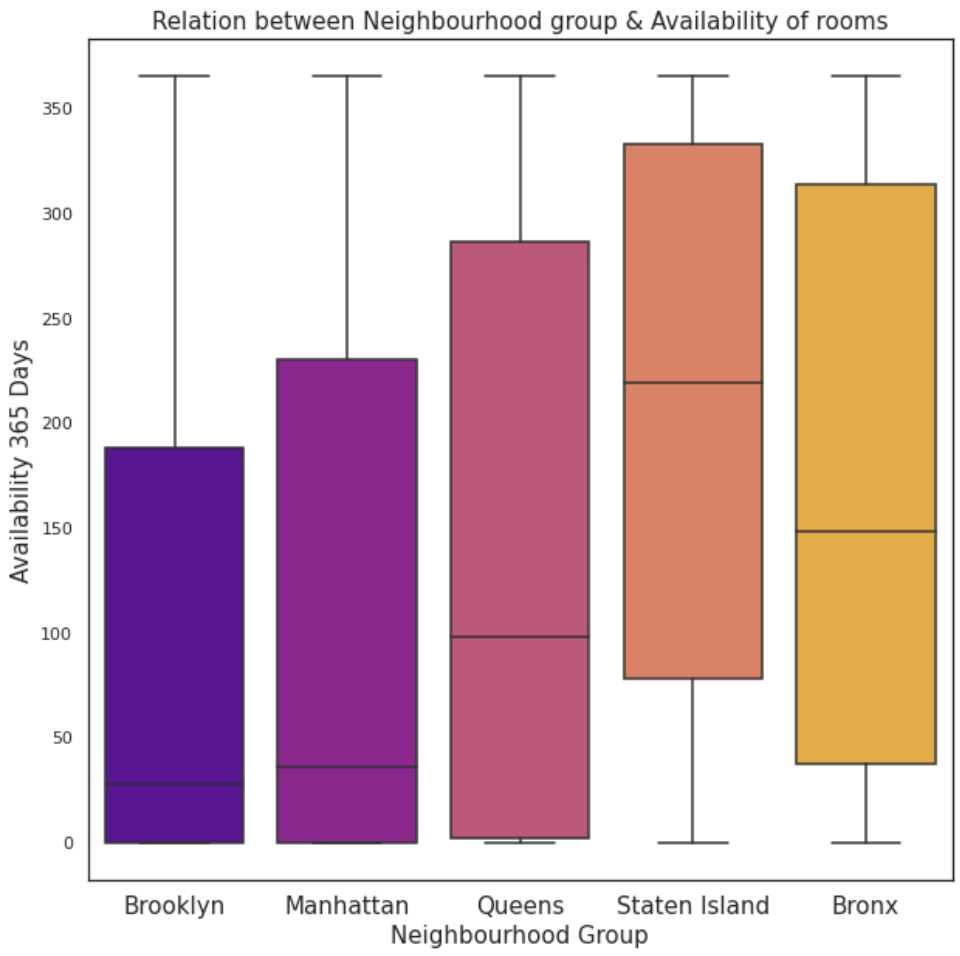
4. very few people preferred to have shared rooms.

5. We will try to catogorise the 'price' like cheep, affordable and expensive and then analyse the 'price' for 'room_type' as per it.

# Area vs Number of reviews :



# Room types and their relation with availability in different neighbourhood groups?(contd.)

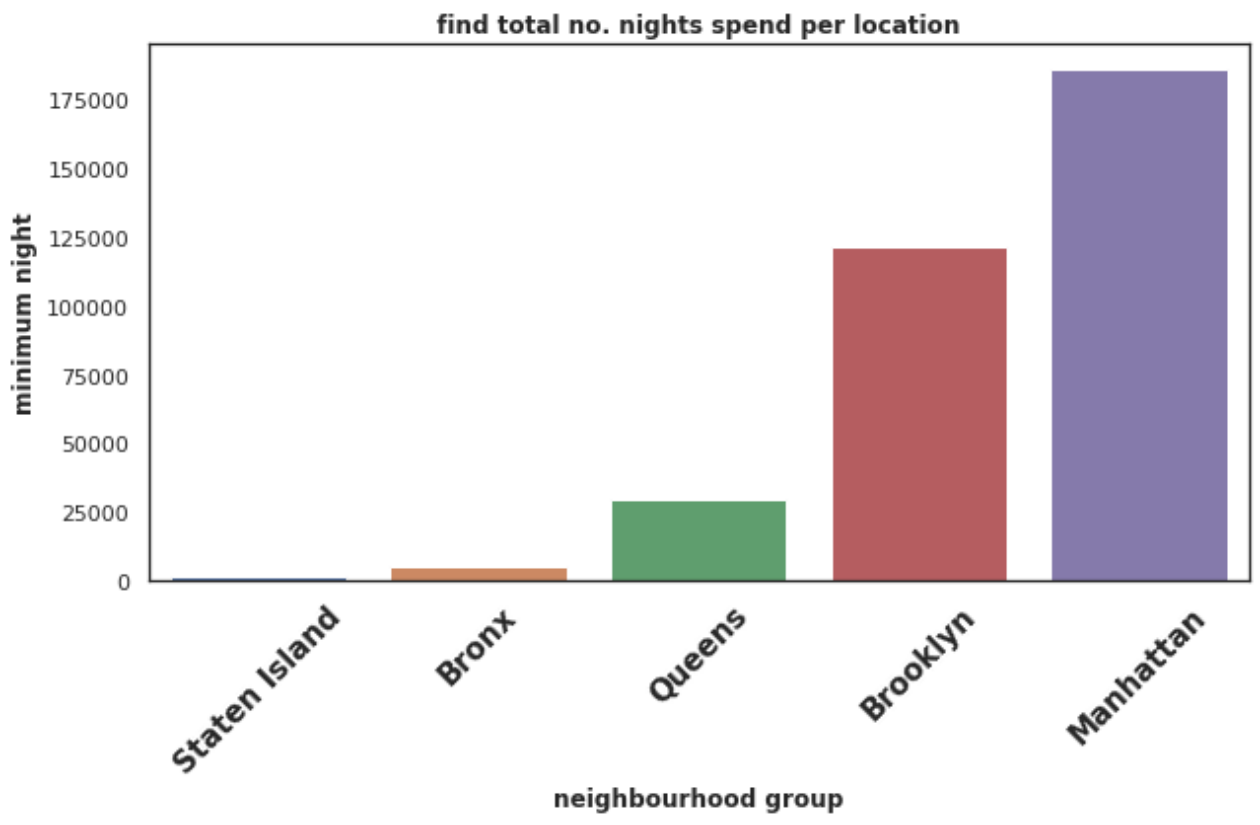



- Brooklyn and Manhattan have the least availability of rooms overall, as low as 0 days.
- Staten Island and Bronx has the highest availability rate overall at around 300 days.
- Form this analysis we can say that people stay for longer duration of time in Private rooms in Brooklyn and Manhattan.

# Price vs Number of Reviews :

# Find total no. nights spend as per location

- The locations are categorized based on its total number of minimum nights spend by customer.
- From here we can get an overall idea of which location the customer prefers more.
- We can state that more customers prefer Manhattan and Brooklyn for night stay as compared to other locations



# Challenges Faced :

- Reading the dataset and understanding the meaning of some columns.
- For answering some of the questions we had to understand
- the business model of airbnb that how they work.
- Handling NaN values,null values and duplicates.
- Designing multiple visualizations to summarize the information in the dataset and successfully communicate the results and trends to the reader

# We defined some points which can help Airbnb in their business :

- Manhattan is the most focused place in New York for hosts to do their business
- Customers pay highest amount in Brooklyn,Queens and Manhattan that is $10,000 and lowest amount is $10.
- For the three types of room type (i.e. Entire home, Shared room, & Private room) average price of entire home is around $157, for Shared room is around $60, and for private room is around $75.
- Top three host base on their turnover are Sonder(nyc),Red awning, Henry and best host is Sonder(nyc

# Conclusion :

We tried to put some light by performed the Extensive EDA for Airbnb dataset as always there is no end to EDA this can be extended in  n-Dimenssions and lots and lots of conclusion can be drawn from EDA this is where 80% of time is been spent by and Data Scientist.

# Thank you